Work probability distribution in single molecule experiments

Alberto Imparato (*) and Luca Peliti (**)

Dipartimento di Scienze Fisiche and Unità INFM, Università "Federico II", Complesso Universitario di Monte S. Angelo, I-80126 Napoli (Italy).

PACS. 87.14.Gg – DNA, RNA.
PACS. 82.37.-j – Single molecule kinetics.
PACS. 82.37.Rs – Single molecule manipulation of proteins and other biological molecules.

Abstract. – We derive and solve a differential equation satisfied by the probability distribution of the work done on a single biomolecule in a mechanical unzipping experiment. The unzipping is described as a thermally activated escape process in an energy landscape. The Jarzynski equality is recovered as an identity, independent of the pulling protocol. This approach allows one to evaluate easily, by numerical integration, the work distribution, once a few parameters of the energy landscape are known.

Introduction. – The introduction of micromanipulation techniques has dramatically improved our knowledge of physical and chemical properties of biological molecules. Such techniques have been used to probe the structure of proteins [1–5] and nucleic acids [6]. A typical experiment consists of pulling the free end of a biomolecule with a controlled force, while its end-to-end distance is measured at the same time. It has been suggested that the study of the kinetics of bond breakage under different loading rates can provide many informations on the molecule internal structure, and in particular allows one to measure the strength of the molecular bonds, and to associate to them a position along the molecular structure [7]. The loading-rate dependent kinetics experiments on biomolecules have been interpreted in terms of thermally activated escape from bond states over a succession of energy barriers, along a one-dimensional energy landscape [8–11].

Usually, because of technical limitations, the molecule pulling process is characterized by time scales much faster than the typical molecular relaxation time. This prevents the possibility to perform the experiment in quasiequilibrium conditions and thus to obtain direct measurements of the thermodynamic state variables. This difficulty can be overcome by exploiting the remarkable equality derived by Jarzynski [12, 13] (Jarzynski equality, JE), and extended by Crooks [14], which allows one to obtain the free energy difference ΔF between two equilibrium states by evaluating the average of $\exp(-\beta W)$, where W is the work performed on the system during the thermodynamic transformation:

$$e^{-\beta \Delta F} = \left\langle e^{-\beta W} \right\rangle . \tag{1}$$

(*) Corresponding author. E-mail: imparato@na.infn.it
(**) Associato INFN, Sezione di Napoli.

Here $\beta = 1/k_B T$, and the average on the rhs is performed over all possible realizations of the process starting from the equilibrium ensemble at a given temperature T. Since this equality holds in general, the information of ΔF can be gathered also by processes so fast that they do not leave the system at equilibrium. On the other hand, the process must be sampled a large number of times in order to obtain a reliable estimate of ΔF, as discussed in [15, 16]. In the same references it is argued that the number of pulling experiments needed to achieve good statistics in the estimate of ΔF increases as the transformation are made more irreversible, for example as the pulling rate is increased. Furthermore the JE gives no information on the probability distribution of work: while the JE magnifies the rare trajectories with $W < \Delta F$, a direct measurements of the probability distribution for these value of W is rather difficult. On the other hand, the effect of such trajectories on the average thermodynamic variables might become significant when the energies involved range between a few and tens of $k_B T$'s, as in the case of molecular bonds.

The aim of this paper is thus to provide an effective method to evaluate the work probability distribution of a molecular pulling process, which can be described as an escape process in an energy landscape, once the main features of the landscape are known. Such a method can be applied independently of the irreversibility of the process, i.e., of the pulling protocol, and for any value of the maximum force.

The article is organized as follows. We first describe our model of escape process. We then introduce a set of differential equations describing the time evolution of the work distribution probability $\phi(W, t)$ and show that the JE follows as an identity. We next consider a simple case of escape process, and discuss the behaviour of the work distribution probability $\phi(W, t)$ obtained for it.

The model. – In a typical unzipping experiment, a force $f(t)$ is applied on one end of the biomolecule, and its elongation $x(t)$ is monitored. Within some limits, $x(t)$ can be considered as a collective coordinate for the system, spanning a one-dimensional (free) energy landscape $E(x)$. This energy landscape will in general be characterized by a set of N minima of energy e_i at position x_i, with $i = 0, \ldots, N-1$, and by a set of $N-1$ maxima of energy E_j at position X_j with $j = 1, \ldots, N-1$, see figure 1. The escape from one minimum e_i over the next maximum E_{i+1} can be viewed as the breaking of a given molecular bond. The evolution of the collective

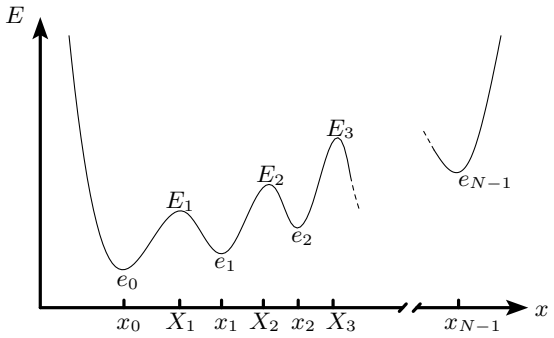

Figure 1 – Generic energy landscape, representing the succession of bonds in a biomolecule.

coordinate will be governed by a stochastic process satisfying detailed balance. If however the variations of the energy E between maxima and minima are large enough with respect to $k_{\mathrm{B}} T$, it is possible to describe this stochastic process as a Markov process with discrete states, corresponding to the extremals of the energy, while the transition probabilities will

be given by the Kramers expression. For simplicity, we shall now consider only transitions between nearest neighbour energy minima $j \to i$, i.e. $i = j \pm 1$. However, we will argue that the results we obtain in the present paper are rather general and can be extended to the case of transitions between any pair of states j and i. Thus the transition rate from one minimum j to a neighboring one i, over the corresponding energy barrier, is given by

$$k_{ij}(t) = \omega_0 \exp\left\{-\beta\left[E_{ij} - e_j - f(t)(X_{ij} - x_j)\right]\right\} , \tag{2}$$

where the energy E_{ij} and the position X_{ij} of the barriers are respectively given by

$$E_{ij} = E_{\max\{i,j\}} , \qquad X_{ij} = X_{\max\{i,j\}} , \tag{3}$$

and where ω_0 is some attempt rate whose value depends on the system characteristics. Let $p_i(t)$ be the probability that the system is in state i at time t: using a Kramers formalism the time evolution of these probabilities is described by the following set of differential equations

$$\frac{\partial p_i(t)}{\partial t} = \sum_j \left[k_{ij}(t)p_j(t) - k_{ji}(t)p_i(t)\right] . \tag{4}$$

The jump process, described by the set of equations (4), is clearly Markovian and it preserves the starting equilibrium ensemble for any fixed value of f. These conditions, together with the assumption that the energies involved in the process are finite, are sufficient for the JE to be recovered for the dynamics described by eqs. (4), as shown by Jarzynski [12, 13] and Crooks [14].

In the following Z_t will denote the the partition function of the canonical ensemble at temperature T and with an applied force equal to $f(t)$:

$$Z_t = \sum_i \exp\left\{-\beta\left[e_i - f(t)x_i\right]\right\} . \tag{5}$$

Work probability distribution. – The probabilities $p_i(t)$ introduced in the previous section, are not sufficient to describe the probability distribution of the work done on the system. We are thus interested in the joint probability distribution $\phi_i(W,t)$ that the system is in state i at time t, while the total work done on it is equal to W.

If the system is in the state i at time t, the work done δW in the time interval δt is given by ([1])

$$\delta W = -\delta t \frac{\partial f(t)}{\partial t} x_i \, . \tag{6}$$

By expanding to first order in δt and δW, the probability distribution $\phi_i(W + \delta W, t + \delta t)$, it is easy to verify that this function satisfies the set of differential Chapman-Kolmogorov equations:

$$\frac{\partial \phi_i(W,t)}{\partial t} = \sum_j \left[k_{ij}(t)\phi_j(W,t) - k_{ji}(t)\phi_i(W,t)\right] + \dot{f}(t)x_i \frac{\partial \phi_i(W,t)}{\partial W} \, . \tag{7}$$

([1])In the present paper we use the ensemble where the force is the externally controlled parameter, while the extension of the molecule x_i fluctuates. This implies that the effective work done on the system, as the force increases of δf, is $\delta W = -x_i \delta f$. In pulling experiments where an optical tweezer is used to pull one of the free ends of the molecule, the actual controlled parameter is the position of the focus of the optical trap. In this situation both the applied force and the molecule extension fluctuate. Here we assume that the fluctuations in $f(t)$ are small, even if the conditions under which this assumption is valid deserve further investigation.

The probability $p_i(t)$ and the probability distribution $\phi_i(W,t)$ are connected by the relation

$$\int dW \phi_i(W,t) = p_i(t) \,, \tag{8}$$

provided that $\phi_i(W,t)$ satisfies the additional boundary conditions $\phi_i(\pm\infty,t) = 0$. By integrating both sides of equation (7) with respect to W, one recovers the set of differential equations (4).

We want now to show that the JE is satisfied for a system characterized by a generic energy landscape as that represented in fig. 1, whose work probability distributions evolve according to eq. (7). Let $\phi_j^l(W,t)$ denote the probability distribution of the work in the state j at time t, when the system initial state is l. They satisfy the set of equations (7) with the initial conditions

$$\phi_j^l(W,t=0) = \begin{cases} \delta(W), & \text{if} \quad j = l\,; \\ 0, & \text{if} \quad j \neq l\,; \end{cases} \tag{9}$$

Thus the average of $e^{-\beta W}$ up to time τ is given by

$$\left\langle e^{-\beta W} \right\rangle_\tau = \sum_l \frac{\exp\left(-\beta e_l\right)}{Z_0} \int e^{-\beta W} \left[\sum_j \phi_j^l(W,\tau) \right] dW = \int e^{-\beta W} \phi(W,\tau) dW \,, \tag{10}$$

where

$$\phi(W,t) = \sum_{lj} \frac{\exp\left(-\beta e_l\right)}{Z_0} \phi_j^l(W,t) \,, \tag{11}$$

is the total work distribution probability. Our goal is to show that the rhs of eq. (10) is equal to $\exp(-\beta \Delta F)$. Let us define the quantity $M_i(t)$ as

$$M_i(t) = \int e^{-\beta W} \sum_l \frac{e^{-\beta e_l}}{Z_0} \phi_i^l(W,t) dW \,, \tag{12}$$

the rhs of eq. (10) can thus be written as

$$\int e^{-\beta W} \phi(W,t) dW = \sum_i M_i(t). \tag{13}$$

By taking the time derivative of $M_i(t)$, and substituting eqs. (7) and (12), we obtain

$$\frac{dM_i(t)}{dt} = \int e^{-\beta W} \sum_l \frac{e^{-\beta e_l}}{Z_0} \frac{\partial}{\partial t} \phi_i^l(W,t) dW \tag{14}$$

$$= \sum_j \left[k_{ij}(t) M_j(t) - k_{ji}(t) M_i(t) \right] + \int e^{-\beta W} \sum_l \frac{e^{-\beta e_l}}{Z_0} \dot{f}(t) x_i \partial_W \phi_i^l(W,t) dW \,. \tag{15}$$

The last term in eq. (15) can be integrated by parts, and using the boundary condition that $\phi_i^l(W,t)$ goes to zero faster than $\exp(\beta W)$ as $W \to -\infty$ (which follows from the existence of the average in eq. (1)), we find that $M_i(t)$ satisfies

$$\frac{dM_i(t)}{dt} = \sum_j \left[k_{ij}(t) M_j(t) - k_{ji}(t) M_i(t) \right] + \beta \dot{f}(t) x_i M_i(t) \,, \tag{16}$$

with the initial conditions $M_i(0) = e^{-\beta e_i}/Z_0$, which follow from the initial value conditions for the functions $\phi_j^l(W,t)$, eq. (9), and from the definition of $M_i(t)$, eq. (12). It is easy to verify that the functions

$$M_i(t) = \frac{\exp\left[-\beta(e_i - f(t)x_i)\right]}{Z_0} \tag{17}$$

are the solutions of the set of equations (16) with the corresponding initial conditions. Substituting this last result into eq. (13), we finally find

$$\int e^{-\beta W}\phi(W,\tau)dW = \sum_i \frac{\exp\left[-\beta(e_i - f(\tau)x_i)\right]}{Z_0} = \frac{Z_\tau}{Z_0}\,, \tag{18}$$

which verifies the JE.

Work distribution probability: a case study. – We consider now a simple system with three minima and two maxima in the energy landscape. The values of the energy minima and maxima, expressed in k_BT units, and their positions, expressed in nm, are $\{e_0 = 0,\, x_0 = 0\}$, $\{E_1 = 10,\, X_1 = 0.6\}$, $\{e_1 = 6,\, x_1 = 0.8\}$, $\{E_2 = 16,\, X_2 = 1.8\}$, $\{e_2 = 12,\, x_2 = 2\}$.

We consider here the case of pulling processes where the force increases linearly with time, i.e. $f = r \cdot t$, where r is the pulling rate. In order to simplify the notation, let

$$\tilde{\phi}_j^l(W,t) = \frac{e^{-\beta e_l}}{Z_0}\phi_j^l(W,t)\,. \tag{19}$$

The work probability distribution $\phi(W,t)$ is thus given by

$$\phi(W,t) = \sum_{l,j} \tilde{\phi}_j^l(W,t)\,. \tag{20}$$

We obtain the probability $\phi(W,t)$, for two values of the pulling rate $r = 1, 10$ pN/s, by solving the set of equations (7) with the parameter choice as indicated above, taking the initial value conditions as given by eq. (9) and the attempt frequency $\omega_0 = 4.4\,\mathrm{s}^{-1}$. With this choice of parameters, we obtain the zero-force transition rate $k_0 = \omega_0 \exp\left[-\beta(E_i - e_{i-1})\right] = 2\times10^{-4}\,\mathrm{s}^{-1}$, in agreement with the zero-force transition rate found in the mechanical unfolding of a simple RNA molecule [6].

The results at different times are shown in figure 2 for $r = 1$ pN/s and $r = 10$ pN/s. In order to analyze the behavior of the tails of the work distribution, the quantity $\phi(W,t)$ for $r = 1$ pN/s is plotted in a log-linear form in figure 3, for the smallest and the largest times considered. Recalling that $\phi(W,t)$ is the sum of the work probability distributions $\tilde{\phi}_j^l(W,t)$ along all the possible trajectories, see eq. (20), in the same figures the main contributions to the total distribution probability of the work are plotted. Inspection of figure 3 suggests that at small times and for small values of $|W|$ the work probability distribution is dominated by those trajectories which start and finish in the state x_0. On the other hand, at any time, at large $|W|$, $\phi(W,t)$ is dominated by those trajectories which finish in the rightmost state x_2, as expected. In particular, in the very large $|W|$ regime, the distribution $\tilde{\phi}_2^2(W,t)$ determines the behavior of $\phi(W,t)$.

Thus, at any time, a given distribution function $\tilde{\phi}_j^l(W,t)$ dominates the behavior of the total work distribution probability $\phi(W,t)$ in a given range of W. In each of such work ranges, the function $\phi(W,t)$ can be well fitted by a gaussian distribution (fits not shown), i.e., the function $\phi(W,t)$ is a superposition of several gaussian functions, each with different mean and variance, rather than a single gaussian function. The gaussian distribution of the work, expected for a slowly perturbed system as discussed in [16,17], is thus not recovered here for the pulling rate $r = 1$ pN/s, which is a lower limit in the pulling experiments of biomolecules.

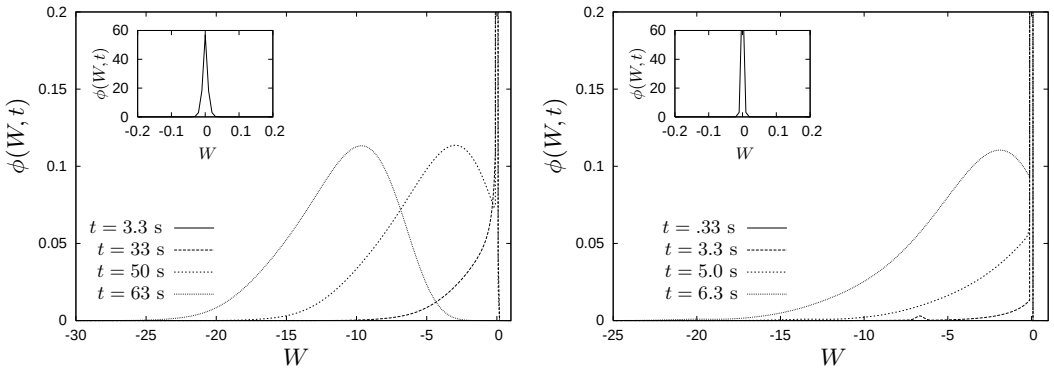

Figure 2 – Plot of the work distribution probability $\phi(W,t)$, as defined by eq. (20) as a function of the work W at different times. Left: $r = 1$ pN/s. Right: $r = 10$ pN/s. Inset: $\phi(W,t)$ at the smallest time here shown (left: $t = 3.3$ s; right: $t = .33$ s).

Discussion. – In the present paper we have introduced a theory which allows us to obtain the work probability distribution performed on a biomolecules during a pulling experiment. The pulling process has been described as a jump process over a succession of energy barriers. This approach allow us to describe the time evolution of the population of the minimal energy states using a set of stochastic differential equations. By discretizing such process, we show that the probability of the discrete trajectories satisfies the Jarzynski equality independently of the pulling protocol. The derivation of the JE within the current formalism can be viewed as the discrete-state-space analogue of the analysis in Sec. II of ref. [13] and in Sec. IV of ref. [18].

We then derive the set of differential equations describing the time evolution of the work probability distributions for each trajectory connecting two arbitrary initial and final states of the system. By summing up these distribution functions over all the initial and final states, we obtain the total work probability distribution $\phi(W,t)$. We show that the JE is recovered for such a distribution function: in this sense, the average over an infinite number of trajectories of eq. (1), is replaced by an average over the probability distribution $\phi(W,t)$ as in eq. (10). In deriving this results, we have assumed that the main features of the energy landscape, i.e., the height and the positions of the wells and of the barriers, are known. For real biomolecules these quantities have to be measured independently for the function $\phi(W,t)$ to be numerically calculated. By considering a very simple energy landscape, we calculate numerically the work probability distribution, and find that it exhibits a non-gaussian behaviour, being rather a superposition of gaussian functions, each corresponding to distinct trajectories between different initial and final state.

We point out that our results still hold if we consider a jump process where transitions between any pair of states are allowed. In fact, as long as the transition rates k_{ij} satisfy the detailed balance condition, eqs. (14,18) are still verified, and thus one recovers the results obtained for the case of jumps between successive states.

* * *

We are grateful to F. Ritort for introducing us to the topic and for interesting discussions.

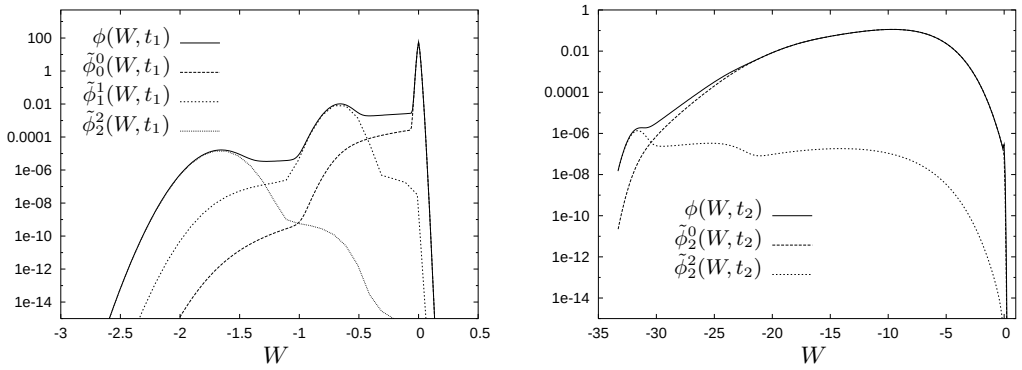

Figure 3 – Log-linear plot of the work distribution probability $\phi(W,t)$ as a function of W for $r = 1$ pN/s, and of the main contributions to the sum on the right hand side of eq. (20), at short times ($t_1 = 3.3$) and at long times ($t_2 = 63$ s).

References

[1] M. RIEF, M. GAUTEL, F. OESTERHELT, J. M. FERNANDEZ and H. E. GAUB, *Science*, **276** (1997) 1109; M. S. KELLERMAYER, S. B. SMITH, H. L. GRANZIER and C. BUSTAMANTE, *Science*, **276** (1997) 1112.

[2] M. G. POIRIER, A. NEMANI, P. GUPTA, S. EROGLU and J. F. MARKO, *Phys. Rev. Lett.*, **86** (2001) 360.

[3] D. A. SIMON, M. STRIGL, M. HOHENADL and R. MERKEL, *Phys. Rev. Lett.*, **83** (1999) 652.

[4] T. NISHIZAKA, R. SEO, H. TADAKUMA, K. KINOSITA and S. ISHIWATA, *Biophys. J.*, **79** (2000) 962.

[5] A. PIERRES, A. M. BENOLIEL, P. BONGRAND and P. A. VAN DER MERWE, *Proc. Natl. Acad. Sci. USA*, **93** (1996) 15114.

[6] J. LIPHARDT, B. ONOA, S.B. SMITH, I. TINOCO and C. BUSTAMANTE, *Science*, **292** (2001) 733; B. ONOA, S. DUMONT, J. LIPHARDT, S.B. SMITH, I. TINOCO and C. BUSTAMANTE, *Science*, **299** (2003) 1892.

[7] E. EVANS and K. RITCHIE, *Biophys. J.*, **72** (1997) 1541; R. MERKEL, P. NASSOY, A. LEUNG, K. RITCHIE and E. EVANS, *Nature*, **397** (1999) 50;

[8] E. EVANS, *Annu. Rev. Biophys. Biomol. Struct.*, **30** (2001) 105.

[9] D. BARTOLO, I. DERÉNYI and A. AJDARI, *Phys. Rev. E*, **65** (2002) 051910; I. DERÉNYI, D. BARTOLO and A. AJDARI, *Biophys. J.*, **86** (2004) 1263.

[10] T. STRUNZ, K. OROSZLAN, I. SCHUMAKOVITCH, H. J. GÜNTHERODT and M. HEGNER, *Biophys. J.*, **79** (2000) 1206.

[11] A. IMPARATO and L. PELITI, *Eur. Phys. J. B*, **39** (2004) 357.

[12] C. JARZYNSKI, *Phys. Rev. Lett.*, **78** (1997) 2690.

[13] C. JARZYNSKI,*Phys. Rev. E*, **56** (1997) 5018.

[14] G.E. CROOKS, *Phys. Rev. E*, **61** (1999) 2361; G.E. CROOKS, *J. Stat. Phys.*, **90** (1998) 1481.

[15] J. LIPHARDT, S. DUMONT, S.B. SMITH, I. TINOCO JR. and C. BUSTAMANTE, *Science*, **296** (2002) 1832.

[16] F.RITORT, C. BUSTAMANTE and I. TINOCO JR., *Proc. Natl. Acad. Sci. USA*, **99** (2002) 13544.

[17] J. GORE, F.RITORT and C. BUSTAMANTE, *Proc. Natl. Acad. Sci. USA*, **100** (2003) 12564.

[18] C. JARZYNSKI, *Dynamics of Dissipation, Lecture Notes in Physics*, edited by P.GARBACZEWSKI AND R.OLKIEWICZ, Vol. **597** (Springer-Verlag, Berlin) 2002, p. 62.